

CLAIMS

What is claimed is:

1. A method of computer based text processing, the method comprising:
 - (a) analyzing selected text units from a digitally coded text file to:
 - i. determine text entities including meaningful N-word groups, phrases, simple sentences and compound sentences;
 - ii. determine the interconnections between the text entities;
 - iii. test the validity of the text entities; and
 - iv. determine a quantitative measure of the significance of each text entity;
 - (b) constructing a multigranular relational text structure incorporating the text entities;and
 - (c) generating output text from the relational text structure.
2. The method of claim 1, wherein the output text is generated from the relational text structure using one or more entity grouping rules.
3. The method of claim 2, wherein the entity grouping rules require that the temporal order of the digitally coded text file be maintained in the output text.
4. The method of claim 2, wherein the entity grouping rules require the grouping of sentences having before action phrases with the same actor.
5. The method of claim 2, wherein the entity grouping rules require the grouping of sentences having after action phrases with the same subject or object upon which the action was executed.
6. The method of claim 2, wherein the entity grouping rules require the grouping of sentences having action phrases with the action belonging to a group of actions represented by the same generalized action.
7. The method of claim 1, wherein the output text generated from the relational text structure conforms to predefined user constraints.

8. The method of claim 7, wherein the predefined user constraints include the volume of output text to be generated, a set of keywords to be reflected in the output text, and the level of generalization of the output text.
9. The method of claim 1, wherein the text file is parsed using a system of natural dividers to recognize text sections, paragraphs, sentences, and words.
10. The method of claim 1, wherein analyzing the text units further comprises comparing each word against one or more dictionaries to determine its part of speech.
11. The method of claim 1, wherein analyzing the text units to determine text entities and test the validity of the text entities comprises attaching each word to the word in its immediate vicinity and using grammatical rules to reduce the ambiguity of the text unit.
12. The method of claim 1, wherein text units are selected from the parsed text file by using windowing and scanning, and analyzing the text units to determine text entities includes searching the text file to find frequent occurrences of words or groups of words.
13. The method of claim 1, wherein analyzing the text units to determine a quantitative measure of the significance is based on the frequency of occurrence, the number of associative links with other text entities, and correspondence of the text unit with predefined user parameters.
14. The method of claim 1, wherein the phrases include action phrases, before action phrases, and after action phrases, and wherein analyzing the text units to determine text entities is accomplished using grammatical rules for action and subject recognition.
15. The method of claim 1, wherein the simple sentences include triplets of interrelated before action phrases, action phrases and after action phrases.

16. The method of claim 1, wherein the output text is generated by grouping before action phrases, action phrases and after action phrases together as a relational structure.

17. The method of claim 1, wherein the volume of output text can be limited to a specified number of words while maintaining a level of text generalization factor (k_G), a depth of descriptive details factor (k_{DD}), and a semantic focus of attention factor (k_{FA}), such that $k_G + k_{DD} + k_{FA} = 1$.

18. A system for providing computer based text processing, the system comprising:

(a) means for analyzing selected text units from a digitally coded parsed text file to:

- i. determine text entities including meaningful N-word groups, phrases, simple sentences and compound sentences;
- ii. determine the interconnections between the text entities;
- iii. test the validity of the text entities; and
- iv. determine a quantitative measure of the significance of each text entity;

(b) means for constructing a multigranular relational text structure incorporating the text entities; and

(c) means for generating output text from the relational text structure.

19. The system of claim 18, wherein the means for generating output text uses one or more entity grouping rules.

20. The system of claim 19, wherein the entity grouping rules require that the temporal order of the digitally coded parsed text file be maintained in the output text.

21. The system of claim 19, wherein the entity grouping rules require the grouping of sentences having before action phrases with the same actor.

22. The system of claim 19, wherein the entity grouping rules require the grouping of sentences having after action phrases with the same subject or object upon which the action was executed.

23. The system of claim 19, wherein the entity grouping rules require the grouping of sentences having action phrases with the action belonging to a group of actions represented by the same generalized action.

24. The system of claim 18, wherein the means for analyzing selected text units attaches each word to the word in its immediate vicinity and uses grammatical rules to reduce the ambiguity of the text unit.

25. The system of claim 24, wherein the predefined user constraints include the volume of output text to be generated, a set of keywords to be reflected in the output text and the level of generalization of the output text.

26. The system of claim 18, wherein the digitally coded parsed text file is parsed using a system of natural dividers to recognize text sections, paragraphs, sentences, and words.

27. The system of claim 18, wherein the means for analyzing selected text units compares each word against one or more dictionaries to determine its part of speech.

28. The system of claim 18, wherein the means for analyzing selected text units uses windowing and scanning to search the text file to find frequent occurrences of words or groups of words.

29. The system of claim 18, wherein the means for analyzing selected text units determines the quantitative measure of the significance of each text entity based on the frequency of occurrence, the number of associative links with other text entities, and the correspondence of the text unit with predefined user parameters.

30. The system of claim 18, wherein phrases include action phrases, before action phrases, and after action phrases, and wherein the means for analyzing selected text units determines text entities using grammatical rules for action and subject recognition.

31. The system of claim 18, wherein simple sentences include triplets of interrelated before action phrases, action phrases and after action phrases.
32. The system of claim 18, wherein the output text generated from the relational text structure conforms to predefined user constraints.
33. The system of claim 18, wherein the means for generating output text groups before action phrases, action phrases and after action phrases together as a relational structure.
34. The system of claim 18, wherein the volume of output text can be limited to a specified number of words while maintaining a level of text generalization factor (k_G), a depth of descriptive details factor (k_{DD}), and a semantic focus of attention factor(k_{FA}), such that $k_G + k_{DD} + k_{FA} = 1$.